

BIOSTATISTICS

Primary Data:

Primary data is a type of data that is collected by researchers directly from main sources through interviews, surveys, experiments, etc. Primary data are usually collected from the source—where the data originally originates from and are regarded as the best kind of data in research.

Primary data is **data that is collected by a researcher from first-hand** sources. There are some methods by which primary data can be collected.

- Creative works (paintings, movie reels, music etc.)
- Diaries,
- Experiments performed by you, the researcher,
- Letters,
- Surveys and censuses,
- Interviews.

A primary source is collected directly from the original source. It is not clouded with someone else's views or judgments.

Secondary Data:-

The term is used in contrast with the term **secondary data**. Secondary data is data gathered from studies, surveys, or experiments that have been run by other people or for other research. Examples of secondary data include:

- Encyclopedias,
- Essays,
- Newspaper opinion pieces,
- Reviews,
- Textbooks.

Typically, a researcher will begin a project by working with secondary data. This allows time to formulate questions and gain an understanding of the issues being dealt with before the more costly and time consuming operation of collecting primary data.

Pie Graph

A **pie Graph** is a type of graph that represents the data in the circular graph. The slices of pie show the relative size of the data. It is a type of pictorial representation of data. A pie chart requires a list of categorical variables and the numerical variables. Here, the term “pie” represents the whole, and the “slices” represent the parts of the whole.

What is a Pie Chart?

The “**pie chart**” is also known as “circle chart”, that divides the circular statistical graphic into sectors or slices in order to illustrate the numerical problems. Each sector

denotes a proportionate part of the whole. To find out the composition of something, Pie-chart works the best at that time.

Formula

The pie chart is an important type of data representation. It contains different segments and sectors in which each segment and sectors of a pie chart forms a certain portion of the total(percentage). The total of all the data is equal to 360°.

The total value of the pie is always 100%.

Therefore, the pie chart formula is given as

$$(\text{Given Data}/\text{Total value of Data}) \times 360^\circ$$

Bar Graph

Bar graphs are the pictorial representation of data. They are also known as bar charts. Bar graphs are one of the means of data handling in statistics.

Bar Graph Definition:

The pictorial representation of a grouped data, in the form of vertical or horizontal rectangular bars, where the lengths of the bars are equivalent to the measure of data, are known as bar graphs or bar charts.

Attributes of Bar graph

The three major attributes of bar graphs are:

- The bar graph helps to compare the different sets of data among different groups easily.
- It shows the relationship using two axes, in which the categories on one axis and the discrete values on the other axis.
- The graph shows the major changes in data over time.

The types of bar charts are as follows:

1. Vertical bar chart
2. Horizontal bar chart

Vertical Bar Graphs

When the grouped data are represented vertically in a graph or chart with the help of bars, where the bars denote the measure of data, such graphs are called vertical bar graphs. The data is represented along the y-axis of the graph, and the height of the bars shows the values.

Horizontal Bar Graphs

When the grouped data are represented horizontally in a chart with the help of bars, then such graphs are called horizontal bar graphs, where the bars show the measure of data. The data is depicted here along the x-axis of the graph, and the length of the bars denote the values.

Properties of Bar Graph

Some of the important properties of a bar graph are as follows:

- All the bars should have a common base.
- Each column in the bar graph should have equal width.
- The height of the bar should correspond to the data value.
- The distance between each bar should be the same.

Advantages and Disadvantages of Bar Chart

Advantages:

- Bar graph summarises the large set of data in simple visual form.
- It displays each category of data in the frequency distribution.
- It clarifies the trend of data better than the table.
- It helps in estimating the key values at a glance.

Disadvantages:

- Sometimes, the bar graph fails to reveal the patterns, cause, effects, etc.
- It can be easily manipulated to yield fake information.

Pictograph

Pictograph Definition

The pictograph is a method to represent the data using images. Each image in the pictograph represents certain things. In other words, pictographs define the frequency of the data using images or symbols, which are relevant to the data. The pictograph is extremely easy to understand, and it is one of the simplest ways to represent the statistical data. In the pictograph, we use a **key**, which denotes the value of the symbol. While using symbols or images, all the symbols should be of the same size.

Advantages of Using Pictographs

Some of the important advantages of using Pictographs are as follows:

- Pictographs are used to express large information in a simple manner.
- It is easy to read, as all the information is provided at one glance.

- It does not require more explanation, as it is universally used.
- It attracts the attention of the viewers or readers, as it has many attractive images

Carto Graph

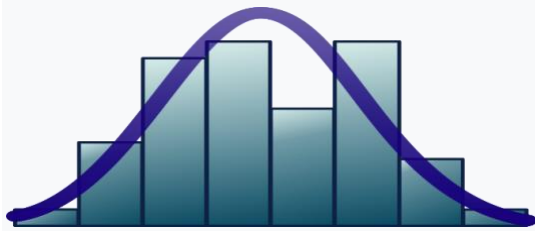
Definition:

The art and science of graphically representing a geographical area, usually on a flat surface such as a map or chart. It may involve the superimposition of political, cultural, or other no geographical divisions onto the representation of a geographical area.

Histo Graph

Definition:

A histogram is used to summarize discrete or continuous data. In other words, it provides a visual interpretation of numerical data by showing the number of data points that fall within a specified range of values (called “bins”). It is similar to a vertical bar graph. However, a histogram, unlike a vertical bar graph, shows no gaps between the bars.



Ogive:

Definition

The Ogive is defined as the frequency distribution graph of a series. The Ogive is a graph of a cumulative distribution, which explains data values on the horizontal plane axis and either the cumulative relative frequencies, the cumulative frequencies or cumulative per cent frequencies on the vertical axis.

Ogive Graph

The graphs of the frequency distribution are frequency graphs that are used to exhibit the characteristics of discrete and continuous data. Such figures are more appealing to the eye than the tabulated data. It helps us to facilitate the comparative study of two or

more frequency distributions. We can relate the shape and pattern of the two frequency distributions.

The two methods of Ogives are:

- Less than Ogive
- Greater than or more than Ogive

Less than Ogive

The frequencies of all preceding classes are added to the frequency of a class. This series is called the less than cumulative series. It is constructed by adding the first-class frequency to the second-class frequency and then to the third class frequency and so on. The downward cumulation results in the less than cumulative series.

Greater than or More than Ogive

The frequencies of the succeeding classes are added to the frequency of a class. This series is called the more than or greater than cumulative series. It is constructed by subtracting the first class, second class frequency from the total, third class frequency from that and so on. The upward cumulation result is greater than or more than the cumulative series.

Uses of Ogive Curve

Ogive Graph or the cumulative frequency graphs are used to find the median of the given set of data. If both, less than and greater than, cumulative frequency curve is drawn on the same graph, we can easily find the median value. The point in which, both the curve intersects, corresponding to the x-axis, gives the median value. Apart from finding the medians, Ogives are used in computing the percentiles of the data set values.

Frequency Polygons

A frequency polygon is a graphical form of representation of data. It is used to depict the shape of the data and to depict trends. It is usually drawn with the help of a histogram but can be drawn without it as well. A histogram is a series of rectangular bars with no space between them and is used to represent frequency distributions.

Definition

A frequency polygon is almost identical to a histogram, which is used to compare sets of data or to display a cumulative frequency distribution. It uses a line graph to represent quantitative data.

IQR:

In descriptive statistics, the **interquartile range (IQR)**, also called the mid spread, middle 50% or H-spread, is a measure of statistical dispersion, being equal to the difference between 75th and 25th percentiles, or between upper and lower quartiles,

$IQR = Q_3 - Q_1$. In other words, the IQR is the third quartile subtracted from the first quartile; these quartiles can be clearly seen on a box plot on the data. It is a trimmed estimator, defined as the 25% trimmed range, and is a commonly used robust measure of scale.

Uses:-

The IQR is used to build box plots, simple graphical representations of a probability distribution.

The IQR is used in businesses as a marker for their income rates.

For a symmetric distribution (where the median equals the mid hinge, the average of the first and third quartiles), half the IQR equals the median absolute deviation (MAD).

The median is the corresponding measure of central tendency.

The IQR can be used to identify outliers

Boxplot And Whisker Plot:

Box plots (also called box-and-whisker plots or box-whisker plots) give a good graphical image of the concentration of the data. They also show how far the extreme values are from most of the data.

A box plot is constructed from five values:

- The first quartile
- The median
- The third quartile
- The maximum value
- The minimum value

We use these values to compare how close other data values are to them.

Purpose of Boxplot

Boxplots are useful tool for getting a sense of central tendency and spread of data. These types of plots are useful diagnostic plots. Use them during initial stages of data analyses. All summary features of box plots are based on ranks (not sums). So, they are less sensitive to extreme values. Box plots reveal asymmetry. Standard deviations are symmetric.

The median splits each batch of numbers in half (center line) The hinges split the remaining halves in half again (the quartiles). Outlier points can be identified, for example, with an asterisk or by id number.

Name :- Asif Ali

Roll. No. 2k21/MLT/ 09